DIVISION OF THE HUMANITIES AND SOCIAL SCIENCES
CALIFORNIA INSTITUTE OF TECHNOLOGY
PASADENA, CALIFORNIA 91125

CORRELATED DISTURBANCES IN DISCRETE CHOICE MODELS: A COMPARISON OF MULTINOMIAL PROBIT MODELS AND LOGIT MODELS

R. Michael Alvarez

Jonathan Nagler
University of California, Riverside
and California Institute of Technology

SOCIAL SCIENCE WORKING PAPER 914

December 1994

CORRELATED DISTURBANCES IN DISCRETE CHOICE MODELS: A COMPARISON OF MULTINOMIAL PROBIT MODELS AND LOGIT MODELS

R. Michael Alvarez Jonathan Nagler

Abstract

In political science, there are many cases where individuals make discrete choices from more than two alternatives. This paper uses Monte Carlo analysis to examine several questions about one class of discrete choice models — those involving both alternative-specific and individual-specific variables on the right-hand side — and demonstrates several findings. First, the use of estimation techniques assuming uncorrelated disturbances across alternatives in discrete choice models can lead to significantly biased parameter estimates. This point is tempered by the observation that probability estimates based on the full choice set generated from such estimates are not likely to be biased enough to lead to incorrect inferences. However, attempts to infer the impact of altering the choice set – such as by removing one of the alternatives – will be less successful. Second, the Generalized Extreme Value (GEV) model is extremely unreliable when the pattern of correlation among the disturbances is not as restricted as the GEV model assumes. GEV estimates may suggest grouping among the choices that is in fact not present in the data. Third, in samples the size of many typical political science applications – 1000 observations – Multinomial Probit (MNP) is capable of recovering precise estimates of the parameters of the systemic component of the model, though MNP is not likely to generate precise estimates of the relationship among the disturbances in samples of this size. Paradoxically, MNP's primary benefit is its ability to uncover relationships among alternatives and to correctly estimate the affect of removing an alternative from the choice set. Thus this paper suggests the increased use of MNP by political scientists examining discrete choice problems *when the central question of interest is the effect of removing an alternative from the choice set.* We demonstrate that for other questions, models positing independent disturbances may be 'close enough.'

CORRELATED DISTURBANCES IN DISCRETE CHOICE MODELS: A COMPARISON OF MULTINOMIAL PROBIT MODELS AND LOGIT MODELS *

R. Michael Alvarez Jonathan Nagler

1 Introduction: Models for Multi-Candidate Elections

In political science many cases arise where individuals choose among more than two alternatives. A prominent set of cases involves candidate choice, and several papers have dealt with this problem (Aldrich and Alvarez 1994, Alvarez and Nagler 1995, Born 1990, Rivers 1988). In these applications, the underlying theoretical assumption is that individuals attempt to maximize expected utility. Thus, for instance, voters see a fixed number of candidates to choose from, and they are assumed to pick the candidate who brings them highest expected utility at the time the choice is being made.

The most commonly used estimation technique for these models has been Multinomial Logit (MNL). However there are substantial drawbacks associated with the use of MNL because it assumes that the disturbances are independent across alternatives. This assumption suggests that if an individual were choosing between three candidates (say, Bush, Perot, and Clinton), then there is no relationship between an individual's disturbance for Bush and his/her disturbance for Clinton or Perot. If individuals view two candidates as having similar attributes in some unmeasured way (say there is some unmeasured characteristic of the candidates, with Clinton and Perot perceived as non-incumbents in an anti-incumbent year) then the disturbances will be correlated. Under this scenario MNL is an inappropriate estimator.

In the MNL framework which contains only individual-specific, as opposed to choice-specific, variables on the right-hand side, this imposes the Independence of Irrelevant

*We thank John Londregan and Langche Zeng for their comments. A previous version of this paper was presented at the Annual Meetings of the American Political Science Association, New York, NY, September 1994. Alvarez thanks the Olin Foundation for support of his research.

Alternatives (IIA) property on the decision-maker. IIA holds when the ratio of the probability of choosing choice j to the probability of choosing choice k is independent of the set of alternatives available. Thus under IIA, adding Perot as an alternative to the choice set of Bush and Clinton would not affect the *relative* probabilities of choosing Bush or Clinton. Hence the econometric assumption of independence inherent in the MNL model leads directly to a substantive assumption of voter behavior.

MNL is an example of a random utility model. In this class of utility models, the individuals' utility is broken into two components: a systemic component and a stochastic component. Thus two models could be based on identical systemic components, yet be distinct because of different stochastic components. In most cases the systemic component contains the substantive parameters of interest. The choice of specification for the stochastic component is frequently a matter of analytic convenience, rather than substance. Note that for IIA to hold a particular specification of *both* the systemic and stochastic component of a random utility model is required. For instance, in the Conditional Logit Model where choice is conditioned on the characteristics of the alternatives, IIA would not be imposed even if the disturbances are assumed to be independent.[1]

Two models which do not assume independent disturbances have been utilized in the econometric literature (but not as extensively in the political science literature). The first of these is the Generalized-Extreme Value model (GEV) originally discussed by McFadden (1973). The GEV model is actually a broad class of discrete choice models, ranging from the MNL model to Nested Logit models (Small 1981). While the GEV model does eliminate the necessity of assuming independence of disturbances, the GEV model requires imposition of specific structure on the grouping of choices. In the three alternative setting two choices must be assumed to be grouped, i.e., viewed as similar alternatives to the decision maker, and the third choice is considered the 'out-choice.' This grouping becomes part of the model specification in GEV. Since we obviously may not know the structure of the choices, often this may even be a focal point of inquiry, this is a severe limitation of the GEV model in political science.

The Multinomial Probit model (MNP) was advanced by Hausman and Wise (1978) as an alternative to the MNL and GEV approaches. The MNP model assumes that the disturbances are distributed multivariate normal. This allows for a very flexible pattern of correlations among disturbances. Thus the MNP model is potentially very useful for political science applications. However, use of MNP has been limited for at least two reasons which have received attention in the literature. First, estimation has been a difficult empirical problem, since the choice probabilities often require the evaluation of high dimensional normal integrals.[2] Some steps have been made to reduce the dimensionality of the integrals needed; others have claimed progress by replacing the choice probabilities with unbiased smooth simulators (McFadden 1989). Both approaches, along with increased availability of computing power, have helped reduce the computational problem posed by MNP.

Second, identification of the elements of the covariance matrix of disturbances has caused problems in prior work. No matter which approach is used to reduce the dimensionality of the integrals involved, there are often many elements in the error covariance matrix to be estimated. For example, in the relatively simple trinomial probit model we have examined in previous work (and continue to examine here), there are three variance terms and three covariance terms which might be estimated in the error covariance matrix. One must be concerned with which of these error covariance coefficients are identified and can be estimated. Beyond the question of technical identification of parameters, MNP models may also exhibit a characteristic dubbed "fragile identification" by Keane (1992). Keane pointed out that even in models where all parameters are technically identified, the nature of the likelihood function makes it difficult in practice to distinguish between the parameters of the systemic component of the model and the estimated covariance elements. Thus one's estimates of all parameters become highly unreliable. The number of elements in the covariance matrix which can be identified depends upon the model specification. In the specification we examine in this paper problems of identification and fragile identification are avoided by the inclusion of a variable that varies across both alternatives and respondents.

Since the computation and identification problems can be overcome, MNP may be a viable and useful technique. However, even though all parameters in our specification are identified, presumably the absence of restrictive assumptions in the MNP model comes at a price: the ability to estimate precisely so many parameters must be limited. The few examples of applied work in political science using MNP (Alvarez and Nagler 1994, 1995) suggest that this problem of obtaining precise estimates may be unresolvable given the size of typical political science data sets. For instance, if three covariance parameters are to be estimated, but in samples of 1000 observations it is impossible to achieve even a 90% level of confidence that the true parameter values are non-zero, then MNP would never allow us to confirm the utility of its use. Alternatively, if MNP allowed for precise estimates of the covariance parameters, but produced standard errors of the estimates of the systemic parameters much greater than MNL or GEV, then MNP might be a useful as a diagnostic tool, but not as an estimation technique.

Thus a major purpose of this paper is to examine the finite sample properties of MNP, and tests its applicability to data sets likely to be encountered in political science. There is little disagreement that MNP has extremely desirable asymptotic properties. But since our data sets are in fact quite finite in size, an estimator with highly desirable theoretical asymptotic properties may not necessarily be an estimator that is of any use to political scientists. There have been few systematic studies of the finite sample properties of the MNP model, and no explicit comparisons of the MNP model to other estimation techniques (Keane 1992; Keane 1994). In fact, in Keane's work on fragile identification (1992) he ignores the sample size question; and he describes estimates generated from a single trial of each model specification. We describe below the full set of questions we address via Monte Carlo analysis.

We proceed by considering six questions. The most obvious question to deal with is whether or not correlations among the disturbances will lead to serious errors of inference if they are ignored in the estimation procedure. If techniques such as MNL which assume independence produce reasonable estimates of the parameters of interest in our models, then we could safely continue to use them. However, if we can demonstrate that techniques assuming independent disturbances cannot produce reliable estimates of the parameters of interest when disturbances are correlated, then there are strong reasons to abandon them for political science applications. Thus our first set of Monte Carlo experiments uses a model assuming independent disturbances (independent probit, i.e.; the MNP model with all correlations between disturbances constrained to be 0) to estimate a model with three alternatives. We estimate the model under six different covariance matrices of the disturbances, and compare the estimates of the parameters of the systemic component to the true parameters. We show that the IP model is liable to generate estimates of substantive parameters such that the expected value of the estimates deviate from the true values by as much as a factor of four. However, given the non-linear nature of these models it is possible that alternative sets of parameters could produce *similar* estimates when probabilities of outcomes are estimated for the same set of independent variables. Thus we present the paradox that the IP parameter-estimates are very far off; but the inferences to be drawn are fairly close to the truth provided the inferences are conditioned on the full choice set used to generate the estimates.

Since GEV has been the most commonly used estimation technique to model correlated disturbances, the second question to consider is how well GEV will enable us to recover parameters of interest if the grouping assumed in the GEV specification is not an accurate reflection of the correlation among disturbances. If the restrictive nature of the assumptions in GEV preclude it from recovering accurate parameter estimates, then GEV would be no more useful than techniques assuming uncorrelated disturbances. We use GEV to estimate our general model on the same six different covariance matrices of the disturbances. In two of the six cases the grouping we estimate is the correct grouping, and in four of the cases it is not. We show that when the specified grouping is incorrect, GEV produces parameter estimates that deviate from the true values by as much as a factor of six. And the results may offer no clue that the specified grouping is incorrect. However, again the estimated effects of each variable on the probability of choosing each of the three alternatives are close to the truth.

The third question we consider is how well MNP recovers estimates of the substantive parameters of interest in samples of 1000 observations. MNP would be an informative estimation technique if it proved to be capable of recovering the substantive parameters of interest with precision, even if it could not recover the correlations among the disturbances. We use MNP to estimate our general model on the same six covariance matrices of the disturbances, and examine the relationship between the estimated systemic parameters and their true values. Even with only 1000 observations MNP produces estimates of the systemic parameters that are generally statistically significant at the 95% level.

The fourth question is whether or not MNP precisely recovers the estimates of the correlation between disturbances. The pattern of correlations among the disturbances may not be of any particular substantive concern in certain political science applications. But having the Monte Carlo results provides a baseline against which to compare results from actual data. If we are able to recover precisely estimates of non-zero correlations in the simulated Monte Carlo data, then we would have greater confidence that failure to recover such parameters in actual data results not from a failing of the estimation technique, but from the fact that non-zero correlations are not present in the process generating our data. MNP turns out to be less successful here, since we are unable to recover significant estimates of the correlations in a majority of relevant cases.

Fifth, it is possible that even if MNP cannot recover precise estimates of the correlations; it may be possible to reject the assumption of independence among the disturbances. A log-likelihood ratio test on all three covariance elements would have more power than the t-tests used on individual coefficient estimates. Thus we compare the log likelihood statistics for MNP and independent probit on 500 trials on identical data. This result is interesting in a perverse way: MNP only allows rejection of independence in approximately half of our cases. Thus there is a high probability that even in the face of correlated disturbances, detection of the correlation is difficult: suggesting that we can *never* have confidence in estimates based on an assumption of uncorrelated disturbances.

Sixth, a question of particular interest in models with multiple choices is what would happen if one additional choice were added, or an existing choice were omitted. We examine whether considerations of the correlation among disturbances offers improved estimates of the effects of omitting one of the three alternatives. Here we find our strongest evidence that the assumption of independence can lead to incorrect inferences, and the strongest case for the use of the MNP model.

In the next section we discuss both the MNP and GEV models in more detail. Following presentation of the models, we describe the Monte Carlo analysis performed and present the results.

2 Multinomial Probit and GEV

2.1 Multinomial Probit

The multinomial probit model allows us to estimate the coefficients of the model without worrying about the implications of the assumption of uncorrelated errors — since we do not have to assume the errors are identically and independently distributed. Instead, we can assume the errors are correlated, and actually estimate these error correlations. Thus MNP is the most flexible random utility model. Here, we present the details of MNP, which follows a framework originally proposed by Hausman and Wise (1978); though we deviate from those authors in the specification of the covariance matrix of the error terms. First, we develop the basics of a multinomial probit model for a three-choice setting. Then we turn to a discussion of modeling the error variances.

We define a random utility function for individual i over each choice j, where $j = 1, 2, 3$:

$$U_{ij} = \overline{U}(X_{ij}, a_i) + \varepsilon_{ij} \tag{1}$$

here X_{ij} is a vector of characteristics unique to choice j relative to decision maker i, a_i is a vector of characteristics unique to the individual decision maker i, ε is a random variable, and $\overline{U}$ defines the systematic component of the utility function of a individual. $\overline{U}$ is assumed to have the following functional form:

$$\overline{U} = \overline{U}(X_{ij}, a_i) = X_{ij}\beta + a_i\psi_j \tag{2}$$

Note that we are assuming that $\overline{U}$ is a linear function of both the characteristics specific to the choice (X_{ij}) and the individual (a_i), with respective parameters β for the choice-specific characteristics and ψ_j for the individual-specific characteristics. The latter coefficient is subscripted by j to indicate that the effects of the individual-specific characteristics vary across choices. Thus there are j ψs for j choices. However, a normalization sets one of those vectors of ψs to zero; and hence $j - 1$ vectors of ψs are actually estimated.

We assume that the random elements of the utility functions, ε_{ij}, have a multivariate normal distribution with mean zero and covariance matrix:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & \\ \sigma_{12} & \sigma_2^2 & \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} \tag{3}$$

Now we assume that the individual chooses the alternative which will bring him or her the greatest utility. This gives the following expression for the probability that the individual would choose the first of the three alternatives:

$$\begin{aligned} P_{i1} &= Pr[(U_{i1} > U_{i2}) \quad \& \quad (U_{i1} > U_{i3})] \\ P_{i1} &= Pr[(\overline{U}_{i1} + \varepsilon_{i1} > \overline{U}_{i2} + \varepsilon_{i2}) \quad \& \quad (\overline{U}_{i1} + \varepsilon_{i1} > \overline{U}_{i3} + \varepsilon_{i3})] \\ P_{i1} &= Pr[(\varepsilon_{i2} - \varepsilon_{i1} < \overline{U}_{i1} - \overline{U}_{i2}) \quad \& \quad (\varepsilon_{i3} - \varepsilon_{i1} < \overline{U}_{i1} - \overline{U}_{i3})] \end{aligned} \tag{4}$$

Since we are actually concerned with differences between disturbances, following Hausman and Wise (1978), we define:

$$\eta_{i,21} = \varepsilon_{i2} - \varepsilon_{i1}, \tag{5}$$

$$\eta_{i,31} = \varepsilon_{i3} - \varepsilon_{i1}. \tag{6}$$

The joint distribution for the $\eta_{i,j1}$ will be bivariate normal, with covariance matrix:

$$\Omega_1 = \begin{bmatrix} \sigma_1^2 + \sigma_2^2 - 2\sigma_{12} & \\ \sigma_1^2 - \sigma_{13} - \sigma_{12} + \sigma_{23} & \sigma_1^2 + \sigma_3^2 - 2\sigma_{13} \end{bmatrix} \tag{7}$$

(8)

This allows us to write the probability that individual i will choose alternative 1 as:

$$P_{i1} = \int_{-\infty}^{\frac{\overline{U}_{i1}-\overline{U}_{i2}}{\sqrt{\sigma_1^2+\sigma_2^2-2\sigma_{12}}}} \int_{-\infty}^{\frac{\overline{U}_{i1}-\overline{U}_{i3}}{\sqrt{\sigma_1^2+\sigma_3^2-2\sigma_{13}}}} b_1(\eta_{21}, \eta_{31}; r_1) d\eta_{21} d\eta_{31} \tag{9}$$

with b_1 being the standardized bivariate normal distribution and r_1 being the correlation between $\eta_{i,21}$ and $\eta_{i,31}$:

$$r_1 = \frac{\sigma_1^2 - \sigma_{13} - \sigma_{12} + \sigma_{23}}{\sqrt{(\sigma_1^2 + \sigma_2^2 - 2\sigma_{12})(\sigma_1^2 + \sigma_3^2 - 2\sigma_{13})}} \tag{10}$$

Similar expressions for P_{i2} and P_{i3} can be easily obtained.

To simplify exposition, we define:

$$\tilde{U}_{i,12} = \frac{\overline{U}_{i1} - \overline{U}_{i2}}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}} \tag{11}$$

which again produce similar definitions for $\tilde{U}_{i,jk}$. This allows us to facilitate writing our earlier expression for P_{i1} as follows:

$$P_{i1} = \int_{-\infty}^{\tilde{U}_{i,12}} \int_{-\infty}^{\tilde{U}_{i,13}} b_1(\eta_{21}, \eta_{31}; r_1) d\eta_{21} d\eta_{31} \tag{12}$$

P_{i2} and P_{i3} can be expressed similarly. Now, the MNP model can readily be estimated by maximum likelihood, once we specify the covariance matrix of the error terms.

Formally, in the MNP model as expressed here in utility differences, there are at most $J(J-1)/2$ free parameters in the covariance matrix which are estimable, where J gives the number of alternatives (Daganzo 1979: p. 95). So, in the trinomial probit case, where $J = 3$, we have *at most* three elements of the covariance matrix which we can estimate.[3] We estimate the three off-diagonal elements of Σ_i and normalize the variances to 1.

2.2 The Generalized-Extreme Value Model

The other model which does not assume the disturbances are correlated is the GEV model.[4] Both the MNP and GEV models depend upon specifying a utility function for each individual where the utility of each alternative consists of a systemic and a stochastic component. It is the stochastic component that differs across the two models. Recall that we specify the utility of the i^{th} individual for the j^{th} choice as follows:

$$U_{ij} = X_{ij}\beta + a_i\psi_j + \epsilon_{ij} \tag{13}$$

where a_i is a vector of characteristics unique to the individual i, X_{ij} is a vector of characteristics unique to alternative j relative to individual i (j = 1,2,3), ψ_j and β are vectors of parameters to be estimated, and ϵ is a disturbance term. For the MNP model we assume that the disturbances have a multivariate normal distribution. The GEV model is derived by assuming that the disturbances have an extreme value distribution, which is given by:

$$F(\epsilon_1, \epsilon_2, \epsilon_3) = exp[-G(e^{-\epsilon_1}, e^{-\epsilon_2}, e^{-\epsilon_3})] \tag{14}$$

where G is a nonnegative function with the constraints that it be homogeneous of degree 1 and always greater than or equal to 0.

Following Maddala (1983) we simplify the notation by letting $Y_{ij} = e^{V_{ij}}$, where V_{ij} denotes the systemic component of utility (i.e., $V_{ij} = a_i\psi_j + X_{ij}\beta$). For cases with three alternatives where the operative hypothesis is that alternatives 2 and 3 are grouped, G can be defined as follows:

$$G(Y_1, Y_2, Y_3) \quad = \quad Y_1 + (Y_2^{1/(1-\sigma)} \; + \; Y_3^{1/(1-\sigma)})^{(1-\sigma)} \tag{15}$$

σ is a parameter to be estimated, and gives a measure of the grouping between alternatives 2 and 3. Furthermore, σ is bounded such that $0 \leq \sigma < 1$.[5] Thus an estimated value of $\hat{\sigma}$ outside these bounds suggests a misspecification problem with the model: the systemic component could be misspecified, or the grouping could be misspecified, or both. This unfortunately leads to the hope that an estimated σ within the range $(0, 1]$ suggests a correct model specification. In fact, as we show below, this is not necessarily the case.

Probabilities are given by:

$$P_{i1} \quad = \quad \frac{Y_{i1}}{G(Y_{i1}, Y_{i2}, Y_{i3})} \tag{16}$$

$$\tag{17}$$

$$P_{i2} \quad = \quad \frac{Y_{i2}^{1/(1-\sigma)}(Y_{i2}^{1/(1-\sigma)} \; + \; Y_{i3}^{1/(1-\sigma)})^{-\sigma}}{G(Y_{i1}, Y_{i2}, Y_{i3})} \tag{18}$$

$$\tag{19}$$

$$P_{i3} \quad = \quad \frac{Y_{i3}^{1/(1-\sigma)}(Y_{i2}^{1/(1-\sigma)} \; + \; Y_{i3}^{1/(1-\sigma)})^{-\sigma}}{G(Y_{i1}, Y_{i2}, Y_{i3})} \tag{20}$$

The log-likelihood function is:

$$LL = \Sigma(y_1 ln(P_{i1}) \; + \; y_2 ln(P_{i2}) \; + \; y_3 ln(P_{i3})) \tag{21}$$

This is straightforward to evaluate with:

$$ln(P_1) \quad = \quad V_1 - ln(G) \tag{22}$$

$$ln(P_2) \quad = \quad \left(\frac{1}{1-\sigma}\right) V_2 \; - \; \sigma ln(e^{V_2^{\frac{1}{1-\sigma}}} \; + \; e^{V_2^{\frac{1}{1-\sigma}}}) \; - \; ln(G) \tag{23}$$

$$ln(P_3) \quad = \quad \left(\frac{1}{1-\sigma}\right) V_3 \; - \; \sigma ln(e^{V_2^{\frac{1}{1-\sigma}}} \; + \; e^{V_2^{\frac{1}{1-\sigma}}}) \; - \; ln(G) \tag{24}$$

2.3 MNP vs. GEV

If two models share similar substantive assumptions then choosing between them would simply be a matter of ease of use and statistical power. Since neither MNP nor GEV make the IIA assumption they do not differ in that strong substantive sense. However, while the models can share the same systemic component and random-utility framework, MNP does allow for a more flexible multivariate distribution of the disturbances.

An obvious advantage of MNP over GEV is that it does not require any a priori grouping of the choices. In specifying the GEV model (for the three alternative case) one needs to assume one of the choices is the 'out-choice' in the construction of the G function. This is of course a substantive assumption, and embodies a prior belief of the structure of the choice process used by individuals. In MNP if such an out-choice exists it will be revealed by the pattern of the covariance estimates. Using GEV is conceptually similar to using MNP, but fixing two of the covariance values to be 0, and only estimating the correlation between the remaining two alternatives. Obviously one could estimate three GEV models, sequentially assuming each of the three choices as the 'out-choice'. For each proposed structure one would want to know if σ falls within the range 0 to 1. This is a backwards way of doing things if the nature of the grouping is a question to be answered, rather than an assumption one is willing to make.

A second disadvantage of GEV becomes clear when we consider grouping structures more complicated than the 'two like alternatives' case. Consider for example that choice 3 shares some unmeasured attributes with choice 1, and shares additional unmeasured attributes with choice 2. This would cause the disturbance of choice 3 to be correlated with the disturbance of choice 1 and of choice 2; though there would be no correlation between the disturbances of choices 1 and 2 (this corresponds to Σ_D below). GEV could not capture this relationship, and MNP would be the appropriate modeling choice.

3 Monte Carlo Tests

In our Monte Carlo simulations, we picked a sample size for each trial of 1000 observations because we think this is typical for survey data: a common case where choice problems arise. The underlying model specification was given by:

$$U_{1i} = .3 - .1 * x_{1i} + 1 * a_{1i} - .2 * a_{2i} + \varepsilon_{1i} \tag{25}$$

$$U_{2i} = .7 - .1 * x_{2i} - .7 * a_{1i} + .04 * a_{2i} + \varepsilon_{2i} \tag{26}$$

$$U_{3i} = .1 * x_{3i} + \varepsilon_{3i} \tag{27}$$

Notice that the "true" model specification here includes two distinct types of coefficients to be estimated. First, we are estimating only one *alternative specific* coefficient, which we denote by β. Furthermore, note that the xs vary by alternative *and* individual. This would correspond to a case such as the ideological distance between a candidate and a voter. Second, we estimate six coefficients of *individual specific* variables, which we will refer to as ψ_{11} thru ψ_{23}. For clarity:

$$U_{1i} = \psi_{11} + \beta * x_{1i} + \psi_{12} * a_{1i} + \psi_{13} * a_{2i} + \varepsilon_{1i} \quad (28)$$
$$U_{2i} = \psi_{21} + \beta * x_{2i} + \psi_{22} * a_{1i} + \psi_{23} * a_{2i} + \varepsilon_{2i} \quad (29)$$
$$U_{3i} = \beta * x_{3i} + \varepsilon_{3i} \quad (30)$$

Our choice of the "true" model specification was again driven by our desire to examine the properties of both MNP and GEV under the conditions most political scientists would face.[6] The *as* were drawn from a normal distribution with mean 0, variance 1; the *xs* were also drawn from a normal distribution with mean 0, variance 1. We estimated this model using six different covariance matrices for the disturbance terms:

$$\Sigma_A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \Sigma_B = \begin{bmatrix} 1 & 0 & .8 \\ 0 & 1 & 0 \\ .8 & 0 & 1 \end{bmatrix}$$

$$\Sigma_C = \begin{bmatrix} 1 & 0 & -.8 \\ 0 & 1 & 0 \\ -.8 & 0 & 1 \end{bmatrix} \qquad \Sigma_D = \begin{bmatrix} 1 & 0 & .5 \\ 0 & 1 & .8 \\ .5 & .8 & 1 \end{bmatrix}$$

$$\Sigma_E = \begin{bmatrix} 1 & 0 & -.5 \\ 0 & 1 & .8 \\ -.5 & .8 & 1 \end{bmatrix} \qquad \Sigma_F = \begin{bmatrix} 1 & -.2 & -.5 \\ -.2 & 1 & .8 \\ -.5 & .8 & 1 \end{bmatrix}$$

For each trial, three error terms corresponding to a particular trivariate normal error covariance matrix were drawn using the algorithm described in Bratley, Fox and Schrage (1983: p. 153-154).[7] In our discussions of the various Monte Carlo simulations below, we refer to these different actual error processes by their respective subscripts, A thru F.

4 Monte Carlo Results

4.1 Independent Probit Results

To provide a baseline for comparison, we begin with six Monte Carlo trials using the Independent Probit model. This is the multinomial probit model, with the covariance terms constrained to be zero so that we maintain the assumption that the errors are homoskedastic and uncorrelated. Thus this would be analogous to the Conditional Logit (CL) model; which is a more general logit model than MNL in that it allows for alternative-specific variables. The relationship between IP and CL is analogous to the relationship between binary probit and binary logit. They assume stochastic components with fundamentally similar properties, with slightly different distributions. In the binary probit and logit case the difference is primarily evident in slightly wider tails for the logistic distribution compared to the normal distribution. Thus any findings for IP should be generalizable to CL.

If in fact a model assuming uncorrelated disturbances produced accurate estimates for the systemic parameters of the model then the utility of *both* GEV and MNP would be called into question. Thus our primary focus here is on the ability of this model to recover the systemic parameters when the disturbances violate the assumptions of the model. Table 1 reports the estimates of the Independent Probit model on 500 trials each of the model with covariance matrices A thru F. We report the mean of the estimated coefficients, the mean estimated standard-error of each coefficient, and the sample standard-error of the estimated coefficient over the 500 trials. Given that these are full-information maximum likelihood estimates, the latter two should be identical. But we report both to verify the reliability of the calculated standard error.

First, if we assume that the errors are homoskedastic and uncorrelated, and the errors correspond to that assumption (column A), the IP model does a very good job reproducing the "true" model parameters. All of the means of the coefficient estimates match the true values to two significant digits. If we assume independent disturbances, and our assumption is incorrect, it is clear that the IP model will lead to inconsistent estimates. In columns B thru F of Table 1, there is substantial error in the estimated model coefficients. For instance, in column B our average estimate of ψ_1 is off by a factor of 2 ($\bar{\hat{\psi}}_{11} = .63$; $\psi_{11} = .3$). Yet the standard error of .10 suggests that we have a very precise estimate of ψ_{11}. In fact, in column B our estimates of every coefficient except β and ψ_{23} are off by more than two standard-errors. And across columns C thru F the situation is equally bad: the estimated coefficients from the model are generally several standard errors away from the true parameters.

However, a quirk of even binary probit models is that the coefficients are not strictly identified. The ratio of the coefficient to the standard error of the disturbances (β/σ) is identified; and we generally normalize σ to be 1 and estimate β/σ. Thus if we are forcing σ to be incorrect here in the IP estimates, it follows that our estimate of β will not yield the true parameters of the model. But, the estimated probabilities, which will be computed by combining $\hat{\beta}$ with the incorrectly constrained $\sigma = 1$ may in fact 'mimic' (to borrow Keane's phrase) the true probabilities.

Thus in Tables 2A and 2B we compute probabilities using the estimated coefficients for the F covariance matrix from MNP and IP. The computed probabilities illustrate the effects of changes in a_1 and x_1 on the predicted probability of choosing each alternative for IP and MNP. These estimates are based on Σ_F. The probabilities in Table 2a are computed with $x_1 = .5$, $x_2 = .5$, $x_3 = -.5$, $a_1 = .5$, $a_2 = -.5$. The table gives the predicted probabilities of choosing alternatives 1, 2, and 3 for such a respondent when $a_1 = .5$; and then indicates the new estimated probability — and corresponding change in probability — as a_1 successively takes the values .25, 0, and -.5. The initial ($a_1 = .5$) IP estimates slightly overpredict the probability of choosing choice 1 and underpredict the probability of choosing choice 3. However, across the entire range of values for a_1 the predicted probabilities for IP and MNP never differ by more than .03.

Table 2B investigates whether the phenomena is also true for changes in the alternative specific variable, x. The probabilities in Table 2B are computed with the same values used in Table 2A, except that x_1 successively takes on the values: 1, .5, 0, and 4. Again, predicted probabilities for MNP and IP never differ by more than .03. Thus the differences between the estimated coefficients and the true parameters do not lead to the errors in prediction one would expect.

Thus the basic conclusion to be drawn from the result in Tables 1 and 2 is that even if there is reason to believe that the errors are correlated (theoretical priors about IIA, a badly specified model, or poorly measured data), estimating a model assuming independent errors may not have very harsh consequences. The parameter estimates generated from such a model are inconsistent, while the standard errors suggest that they are very accurate estimates. The estimated parameters combined with incorrectly constrained sigmas do in fact appear to mimic the true probabilities, provided they are computed over the full choice set.[8] However, the estimated effects of changes in independent variables are much more accurate than one would guess looking at the parameter estimates. And while we do not report alternative logit models here which assume uncorrelated disturbances, there is no reason to expect them to perform differently than IP.

4.2 GEV Results

We next examine the performance of GEV under the same circumstances. We estimated a model with the same systemic component, but assumed that the disturbances have a generalized extreme value distribution, similar to equation (15), except that choices 1 and 3 grouped. This should produce parameter estimates under covariance matrices A, B, and C that are close to the true values. Since the disturbances in the trials are actually multivariate normal, not extreme-value, our estimates will not be exact.[9] However, provided the pattern of correlations across disturbances is as we postulated, we do not expect the differences to be any worse than what we would experience runing logit instead of probit if the disturbances in a binary choice problem were really normal, or vice-versa.[10]

The GEV model performs quite well when in fact the disturbances are independent (column A). All estimates are within less than one-half standard error of the true values. GEV also does well when the postulated grouping accurately reflects the structure of the covariance matrix of the disturbances. In column B the covariance matrix has a positive correlation between choices 1 and 3, and the remaining two off-diagonal elements are 0. This corresponds to the grouping of choices 1 and 3. Estimated coefficients are within a standard error of the true parameters. And our estimate of σ for the grouping is .64; the square of the actual correlation between the disturbances for choices 1 and 3. For the case where the correlation between the disturbances of choice 1 and 3 is negative (column C), GEV does not do as well. The estimates of the coefficients are still within 1 standard-error of the actual values, but none are as accurate as the previous case. The estimated value of σ is still close to the negative square of the correlation; but the standard error is much larger. The estimates are reported in Table 3.

In each of the cases in columns D thru F the covariance matrix of the disturbances contains more than 1 non-zero off-diagonal element. Thus GEV is incapable of capturing the relationships between the disturbances. As Table 3 indicates, in such cases GEV does a very poor job of recovering the systemic parameters. The estimates of ψ_{11} are between 4 and 6 times too large. Other estimates are similarly bad: ψ_{21}, ψ_{22}, and ψ_{23} are off approximately by factors of 3. Thus again, a model assuming an incorrect relationship between disturbances produces systemic estimates that are not biased by a small amount, but are biased by a *large* amount.

What is also problematic with the GEV estimates is the estimate of the grouping parameter σ in column D. Here the estimate of σ is .59, with a t-statistic of 1.64. Thus if one used the criteria of estimating σ to be within (0,1), this model would suggest that the postulated grouping of choices 1 and 3 is correct. However, in covariance matrix D the grouping is actually much more complex. In fact the highest correlation between disturbances is between choice 2 and choice 3, not choices 1 and 3. Thus a researcher estimating case D using GEV would be tempted to infer that choices 1 and 3 are grouped, when in fact this is not the case. Furthermore, estimates of σ in case E and F are -.49 and -.26, respectively, and reach statistical significance. In such cases the researcher would traditionally be unable to reject independence, and presumably resort to IP or CL. However, both these models would perform badly.

The interpretation of σ in GEV estimates is generally not very rigorous. We emphasize here that an estimate of σ outside the interval (0,1] does *not* suggest abandoning the search for the correct grouping and assuming independence. Rather, it suggests considering an estimation procedure capable of specifying more general grouping than GEV can handle. Similarly, we emphasize that an estimate of σ within (0,1] does *not* guarantee that the correct grouping of alternatives has been identified. We turn below to the Multinomial Probit Model as an example of an estimation technique which can handle more complex structures of the disturbances.

However, again it is important to see if the GEV model produces accurate estimates of probabilities even when the parameter estimates are not accurate. Table 3A gives estimated probabilities using the GEV model for different values of the individual-specific variable a. No estimated probability differs from the corresponding probabilities generated by the MNP estimates by more than .04. Table 3B gives probabilities estimated by GEV different values of the alternative-specific variable x. Again, no estimated probability differs from probabilities computed from MNP estimates by more than .04. Thus, as with IP estimates, bad GEV estimates appear capable of producing good predictions.

4.3 Effects of Removing an Alternative

Since we find it inconceivable that estimates as bad as those generated by IP and GEV can really be as good as MNP estimates, we put them to one more test. Rather than predicting the probability of choosing each of the three alternatives, we use them to produce the probability of choosing alternative 1 when alternative 3 is removed from the choice set. This *should* cause problems for the IP and GEV estimates. After all, the disturbance for choice 3 is negatively correlated with the disturbance for choice 1 (σ_{13} = -.5) and positively correlated with the disturbance for choice 2 (σ_{23} = .8). Thus if choice 3 is eliminated it should *help* choice 2 more than IP realizes. In other words, IP should underpredict choice 2 and overpredict choice 1. This is exactly what happens. Table 4 gives the estimated probability for choosing alternative 1 when alternative 3 is eliminated from the choice set generated by the MNP, IP, and GEV estimates. IP overpredicts this probability by .06, and GEV overpredicts by .07. For the alternative-specific variables, both IP and GEV both overpredict the estimated changes in probability, but not by as large a margin (.02 and .04, respectively). The important point contained in Table 4 is that the primary problem with the IP and GEV models are that they will produce incorrect predictions about the choices made by decision makers when the choice set is altered.

4.4 Multinomial Probit

Next we present our Monte Carlo estimates using the Multinomial Probit model. Since multinomial probit is the 'correct' model for the data, we should recover consistent estimates of the parameters of interest. Hence our primary concern here is our ability to recover precise estimates of these parameters with samples of 1000 observations. We are concerned both with recovering the systemic parameters, and with recovering estimates of the correlation across the disturbances. We estimated a multinomial probit model for each of our six covariance matrices (A thru F) in which we constrained the diagonal elements to be 1, and estimated the three off-diagonal elements. This is the most general specification available to us. The estimates of the systemic coefficients are reported in Table 5.

First, we do accurately recover the systemic parameters. Four of the estimates of β are within 10% of the actual parameter, and all are within a standard error. Similarly, five of the six estimates of ψ_{11} are within 20% of the actual parameter, and all are well within one-half standard error. In fact all 42 coefficient estimates in Table 5 are within one standard error of the actual parameter value. Thus MNP can be a viable estimation technique for cases with only 1000 observations. MNP can recover significant estimates of the systemic parameters, which are generally the parameters of interest.

It is informative to compare the standard errors here with the standard errors for the independent probit model for column A. MNP is generally able to recover estimates that are significant at traditional levels. For the parameters β, ψ_{12}, ψ_{13}, ψ_{21}, and ψ_{22} the MNP estimates are on average more than twice their standard errors for at least five of the six covariance matrices examined. In this case the independent model is appropriate, and in Table 5 we are actually attempting to estimate values for 3 correlations that are in fact 0. Here MNP produces standard errors that are as much as four times as large as the standard errors from the IP model, though on average they appear to be about twice as large. On the one hand this is quite a price to pay to allow for a more general model structure. On the other hand, considering how bad the IP estimates can be in the face of correlated disturbances, this is a very small price (or perhaps a very necessary price) to pay.

However, it remains open as to whether MNP can recover significant estimates of the elements of the covariance matrix of the disturbances. There are three reasons we might want to recover these parameters. First, they could be of substantive interest should we have beliefs about the grouping process of decision-makers. Second, we might want to test for independence of the disturbances as a diagnostic tool. Third, they help us understand the effects of removing an alternative from the choice set. Table 6 below shows the estimated covariance matrices from our Monte Carlo experiments, as well as the actual covariance matrices used to produce the data. In the six separate experiments we try to recover a total of nine non-zero correlations. We are able to recover estimates significantly different from zero by normal statistical criteria in only three of these cases. Thus the MNP model can recover three error covariances, but the precision of the estimated covariances in samples of the size political scientists are likely to encounter is not necessarily likely to allow us to determine the nature of the grouping.

However, even if the MNP model does not allow for precise estimates of *each* element of the covariance matrix, it would be useful if MNP could be used to examine if *any* off-diagonal element of the covariance matrix is non-zero. This would be a test of independence of the disturbances. Such a test would be extremely useful as a diagnostic tool. We have seen how badly estimation techniques assuming uncorrelated disturbances perform when the disturbances are correlated. Having a statistical test available that could suggest to us when independence is a viable assumption would allow us to continue to use estimation techniques such as CL and IP under appropriate circumstances, since we would be able to identify such circumstances. Such a statistical test is readily available via a log-likelihood ratio test using the MNP and IP models. By estimating the MNP model, then constraining the covariance elements to be zero and estimating IP, twice the difference between the log-likelihood values will be distributed χ^2 with three degrees of freedom.

We performed this test for the 500 trials with Σ_E. In 48% of the trials the χ^2 value was large enough to reject independence at the 90% confidence level. In 32% of the trials the computed χ^2 value was large enough to reject independence at the 95% confidence level. This is not too encouraging towards MNP's utility as a diagnostic tool. Remember that Σ_E is a covariance matrix with large correlations among disturbance terms. Thus even in as 'messy' a case as this, one would reject independence only half the time. However, paradoxically, it is the weakness of MNP as a diagnostic tool that suggests even more strongly the importance of its use as an estimation technique. For these tests suggest that even in the face of highly correlated disturbances, which we have shown cause significant bias in IP estimates of the substantive parameters, we will not be able to explicitly demonstrate that the disturbances are correlated. This means that in all cases, estimates generated via IP will be suspect.

5 Discussion

This paper has two main purposes. One main purpose of this paper was to examine the small sample performance of the MNP model to determine its potential usefulness for political science applications. Our other main purpose was to determine the consequences of *not* running MNP and instead relying on a technique assuming independent disturbances. We have very clearly shown that when the errors are correlated, Independent Probit does a poor job of producing estimates of the parameters of the systemic component of the model. But we have just as clearly shown that Independent Probit will produce reasonably faithful estimates of probabilities, and hence lead to accurate inferences, *provided* these are based on the full choice set. Our examination of GEV has shown that GEV has a similar property to IP: it tends to produce estimates of the systemic parameters that are badly biased, but yield reasonably good estimates of the effects of changes in probabilities based on changes in independent variables. However, GEV may give results suggesting that an incorrect grouping is a correct model specification. Conversely, as would be expected, even if the errors are uncorrelated but we attempt to estimate them, the MNP model produces consistent estimates of the parameters of the model. The failing we have demonstrated in both IP and GEV is in how they perform in predicting probabilities when the choice set is altered.

We cannot emphasize enough that these conclusions are all based on the formulation of the systemic component of the model that we have adopted. In particular, we are using a very general form of the systemic model here: one including both individual specific variables and alternative-specific variables. It is important to realize that the more commonly used Multinomial Logit Model does not include alternative-specific variables. In cases where MNL *may* be the appropriate model (i.e., where the correct model does not include alternative specific variables), correlations among the disturbances may or may not lead to results similar to those reported here. Research on that specific question is beyond the scope of this paper.

The price one pays for running MNP rather than IP (or a logit formulation) is that the standard errors of the coefficient estimates are larger than they would be under IP. However, we have shown that with samples of 1000 observations it is possible to recover statistically significant estimates of the substantive parameters of the model. This holds for the two types of coefficients we have estimated here: alternative-specific coefficients as well as individual-specific coefficients. However, in applications where understanding the error covariances is important, MNP may not be useful to political scientists with data sets on the order of 1000 observations. The problem with MNP as we see it now is that it is not very likely that in samples of 1000 observations MNP will produce precise estimates of the correlations among the disturbances. While this is troubling, it should not lead us to abandon the MNP technique. The fact that correlations between disturbances are difficult to precisely estimate should not lead us to adopt model specifications and estimation techniques that assume those correlations are zero. What our results show is that we should be aware that the MNP model will give us consistent estimates of the model parameters, but we are unlikely to have a reliable idea of what the error covariance terms are.

So, our practical advice is that when dealing with multiple, discrete, and unordered choice data whether one is willing to assume that the error terms are uncorrelated should depend upon what one wishes to learn. It appears that basic inferences of the effects of independent variables on the probability of choosing each alternative will be correct even if this assumption is maintained when it is in fact false. However, if the researcher is interested in learning of the effects of omitting a choice then the independence assumption should be avoided. Instead, we recommend the use of the MNP model: it provides a means to obtain consistent estimates of the parameters of the systemic component of our models *and* correctly computes probabilities based on alternative choice sets.

Increased availability of sophisticated statistical software and powerful computing capability is bringing a range of models within reach of use by political scientists. However, before rushing to embrace these models the discipline needs to examine how well suited they are to the types of data sets likely to be encountered. We have demonstrated here that applying CL and GEV to a particular set of questions would be a bad idea for the discipline. And we have shown that MNP estimation is practical on data sets of the size likely to be available, but that samples of more than 1000 observations will probably be necessary to precisely learn the relationship among disturbances.

The research we have presented here is of course not an exhaustive treatment of MNP or GEV. As empirical researchers are well aware, one only learns the properties of estimation techniques through extensive use. And researchers may well find cases with 1000 observations where MNP produces extremely precise estimates of correlations among disturbances. By providing evidence for the benefits of MNP in political science over GEV and CL we hope to encourage researchers to begin to accumulate such extensive experience. We think choice models are in their infancy in political science. In this paper we have tried to lay a piece of the groundwork for steps toward estimating richer models.

Table 1

Estimates of Independent Probit Across Hypothetical Error Structures

	Error Covariance Matrix: True	A	B	C	D	E	F
β	-.1	-.10	-.13	-.09	-.12	-.14	-.13[a]
		.03	.04	.03	.04	.04	.04[b]
		.03	.03	.03	.04	.03	.03[c]
ψ_{11}	.3	.30	.63	.21	1.99	.90	1.03
		.08	.10	.08	.16	.09	.10
		.08	.09	.07	.15	.09	.09
ψ_{12}	1	1.0	1.79	.79	.68	.31	.26
		.10	.13	.09	.18	.10	.10
		.10	.13	.09	.19	.10	.10
ψ_{13}	-.2	-.20	-.37	-.16	-.23	-.11	-.11
		.08	.09	.08	.14	.08	.08
		.07	.08	.06	.14	.07	.07
ψ_{21}	.7	.70	1.27	.44	2.38	1.07	1.21
		.07	.09	.07	.16	.09	.10
		.07	.08	.07	.14	.09	.09
ψ_{22}	-.7	-.70	-.22	-.91	-1.06	-1.75	-1.55
		.09	.09	.09	.18	.12	.12
		.09	.10	.09	.19	.13	.12
ψ_{23}	.04	.04	-.06	.06	.004	.14	.12
		.08	.08	.07	.15	.09	.09
		.06	.07	.06	.14	.07	.08

[a] The first row gives the average of our estimate for 500 trials.

[b] The second row gives the sample standard deviation of our 500 trials.

[c] The third row gives the average our 500 trials of the standard error computed by our maximum likelihood code.

Table 2A

Predicted Changes in Probability for Changes in Individual Specific Variable

MNP Estimates

	Chc 1	Δ	Chc 2	Δ	Chc 3	Δ
a_1 = .5	.58	–	.22	–	.19	–
a_1 = .25	.49	(.09)	.33	(-.11)	.18	(.02)
a_1 = 0	.40	(.19)	.46	(-.23)	.15	(.05)
a_1 = -.5	.22	(.36)	.70	(-.48)	.08	(.11)

IP Estimates

	Chc 1	Δ	Chc 2	Δ	Chc 3	Δ
a_1 = .5	.60	–	.24	–	.16 –	
a_1 = .25	.51	(.09)	.35	(-.11)	.14	(.02)
a_1 = 0	.40	(.20)	.48	(-.24)	.12	(.04)
a_1 = -.5	.21	(.39)	.73	(-.49)	.07	(.09)

Table 2B

Predicted Changes in Probability for Changes in Alternative Specific Variable

MNP Estimates

	Chc 1	Δ	Chc 2	Δ	Chc 3	Δ
$x_1 = 1$	.58	–	.22	–	.19	–
$x_1 = .5$	.60	(-.01)	.22	(.01)	.19	(.01)
$x_1 = 0$	.61	(-.03)	.21	(.01)	.18	(.01)
$x_1 = 4$	.50	(.08)	.27	(-.05)	.23	(-.04)

IP Estimates

	Chc 1	Δ	Chc 2	Δ	Chc 3	Δ
$x_1 = 1$	.60	–	.24	–	.16	–
$x_1 = .5$	.62	(-.02)	.22	(.01)	.16	(.01)
$x_1 = 0$	.64	(-.04)	.21	(.02)	.15	(.02)
$x_1 = 4$	.48	(.12)	.30	(-.07)	.21	(-.05)

Table 3

GEV Estimates With Y_1 and Y_3 As Grouped Alternatives

	True	A	B	C	D	E	F
β	-.1	-.13	-.12	-.14	-.13	-.20	-.17
		.05	.04	.06	.06	.07	.06
		.05	.04	.05	.06	.06	.06
ψ_{11}	.3	.34	.33	.37	1.31	1.80	1.77
		.20	.13	.26	1.11	1.11	1.16
		.19	.12	.26	1.14	1.28	1.39
ψ_{12}	1	1.33	1.15	1.62	.30	.42	.26
		.58	.39	.79	.31	.39	.27
		.51	.35	.68	.30	.38	.29
ψ_{13}	-.2	-.27	-.23	-.34	-.15	-.21	-.17
		.16	.10	.23	.22	.24	.20
		.14	.09	.20	.19	.22	.20
ψ_{21}	.7	.88	.79	1.07	1.77	2.14	2.05
		.46	.23	.71	1.13	1.35	1.33
		.41	.21	.66	1.17	1.56	1.61
ψ_{22}	-.7	-.87	-.88	-.88	-1.81	-2.28	-2.06
		.31	.25	.40	.31	.30	.24
		.28	.22	.36	.31	.33	.27
ψ_{23}	.04	.04	.05	.02	.13	.13	.12
		.12	.10	.14	.23	.20	.18
		.11	.09	.13	.20	.19	.18
σ		-.004	.64	-.63	.59	-.49	-.26
		.47	.14	.81	.36	.94	.80
		.41	.12	.74	.36	1.06	.99

[a] The first row gives the average of our estimate.

[b] The second row gives the sample standard deviation.

[c] The third row gives the average of the standard error computed by our maximum likelihood code.

Table 3A

GEV: Predicted Changes in Probability for Changes in Individual Specific Variables

	Chc 1	Δ	Chc 2	Δ	Chc 3	Δ
$a_1 = .5$	.62	–	.23	–	.16	–
$a_1 = .25$	.52	(.10)	.34	(-.11)	.14	(.02)
$a_1 = 0$	.41	(.21)	.48	(-.25)	.11	(.04)
$a_1 = -.5$	.20	(.42)	.74	(-.51)	.06	(.09)

Table 3B

GEV: Predicted Changes in Probability for Changes in Alternative Specific Variables

	Chc 1	Δ	Chc 2	Δ	Chc 3	Δ
$x_1 = 1$	.62	–	.23	–	.16	–
$x_1 = .5$	.63	(-.02)	.22	(.01)	.15	(.01)
$x_1 = 0$	.65	(-.04)	.20	(.02)	.14	(.01)
$x_1 = 4$	.50	(.11)	.30	(-.08)	.19	(-.04)

Table 4

Predicted Probability and Changes for Choosing Alternative One When Alternative Three is Removed from the Choice Set

Changes in Individual Specific Variable	MNP		IP		GEV	
	$\hat{P}_1$	Δ	$\hat{P}_1$	Δ	$\hat{P}_1$	Δ
$a_1 = .5$	.65	–	71	–	72	–
$a_1 = .25$	54	(.11)	59	(.12)	59	(.13)
$a_1 = 0$	.43	(.22)	.46	(.25)	.44	(.27)
$a_1 =$ -.5	.23	(.42)	.23	(.48)	.20	(.52)

Changes in Alternative Specific Variable						
	$\hat{P}_1$	Δ	$\hat{P}_2$	Δ	$\hat{P}_3$	Δ
$x_1 = 1$	.65	–	.71	–	.72	–
$x_1 = .5$	.66	(-.01)	.72	(-.02)	.74	(-.02)
$x_1 = 0$	.68	(-.03)	.74	(-.03)	.75	(-.03)
$x_1 = 4$	.57	(.08)	.61	(.10)	.60	(.12)

Table 5

Estimates of Coefficients via Multinomial Probit: Three Elements of the Covariance Matrix Estimated

		Error Covariance Matrix:					
	True	A	B	C	D	E	F
β	-.1	-.09	-.11	-.08	-.13	-.11	-.11[a]
		.03	.03	.03	.04	.04	.04[b]
ψ_{11}	.3	.24	.27	.24	.65	.29	.34
		.39	.37	.34	.56	.33	.36
ψ_{12}	1	.94	1.18	.78	1.01	.89	.83
		.39	.41	.35	.42	.29	.29
ψ_{13}	-.2	-.18	-.23	-.16	-.22	-.18	-.17
		.09	.09	.09	.10	.09	.09
ψ_{21}	.7	.62	.73	.52	1.11	.65	.68
		.27	.28	.24	.52	.29	.32
ψ_{22}	-.7	-.65	-.70	-.69	-.96	-.96	-.91
		.31	.22	.34	.43	.41	.39
ψ_{23}	.04	.04	.03	.04	.06	.07	.06
		.08	.08	.07	.11	.07	.07

[a] The first row gives the average of our estimate for 500 trials.

[b] The second row gives the sample standard deviation of our 500 trials.

Table 6

Estimates of Correlations Between Disturbances: Recovering Three Correlations

	Actual Correlations			Estimated Correlations		
	σ_{12}	σ_{13}	σ_{23}	$\hat{\sigma}_{12}$	$\hat{\sigma}_{13}$	$\hat{\sigma}_{23}$
Set A	0	0	0	.05	.07	.08[a]
				.28	.42	.47 [b]
				.29	.41	.46[c]
Set B	0	.8	0	-.19	.67	.08
				.24	.20	.36
				.25	.20	.38
Set C	0	-.8	0	.18	-.16	.13
				.27	.45	.47
				.29	.42	.45
Set D	0	.5	.8	-.33	.27	.49
				.22	.41	.33
				.26	.44	.37
Set E	0	-.5	.8	.09	-.27	.68
				.24	.38	.22
				.22	.32	.20
Set F	-.2	-.5	.8	-.17	-.20	.66
				.24	.38	.23
				.26	.39	.23

[a] The first row gives the average of our estimate for 500 trials.

[b] The second row gives the sample standard deviation of our 500 trials.

[c] The third row gives the average our 500 trials of the standard error computed by our maximum likelihood code.

6 Endnotes

[1]However, the independence of the disturbances would still generate biased estimates of the parameters of interest for Conditional Logit if the errors were in fact correlated.

[2]Typically, the number of integral dimensions is roughly equivalent to the number of alternatives. In the trinomial probit model, for example, with three choices there would normally be three dimensions to the multivariate normal integral. However, following the lead of Hausman and Wise (1978), by examining not the utility functions but rather the differences between utility functions, the number of dimensions can be reduced to two. Generally this has been shown to work for up to five alternatives.

[3]Daganzo also states that as a regularity condition of the MNP model, the specified error covariance matrix should be positive definite across all possible values of the latent parameters (1979). This has led some to argue that error covariance matrices should be estimated as a practical matter as Cholesky factorizations of the original error covariance matrix (e.g. Bunch 1991). This will insure that the regularity conditions are generally met for a formally-identified covariance matrix. Additionally, expressing the error covariance matrix as a Cholesky factorization also allows the modeler to examine all possible transformations between the Cholesky factorization and the original error covariance matrix to insure their validity.

[4]GEV is equivalent to the general specification of nested logit; however one can run a model referred to as nested logit which does impose IIA. In such a case the coefficient of the 'inclusive value' is constrained to be one. See Maddala (1983).

[5]GEV is equivalent to the Nested Multinomial Logit (NMNL) model where the coefficient of σ in NMNL is not constrained to be 1. It has been shown that NMNL can be estimated sequentially using binary logit. One problem with this technique is that it does not produce correct standard errors. Amemiya (1978) provides a correction for the covariance matrix of the estimates. But it seems much more straightforward to adopt an estimation technique directly producing an asymptotically consistent covariance matrix than to produce a covariance matrix needing correction.

Another potential problem with the sequential estimation technique is that it may not be as flexible in allowing for the formulation of the systemic component of the utility function. The GEV model is quite robust in this respect. All estimates presented here are full information maximum likelihood estimates of the GEV model.

[6]In fact, in our previous work with the MNP and GEV models on data taken from both U.S. and U.K. national elections, we have estimated models very similar to this. In future Monte Carlo simulations, it will be very important to replicate all of our Monte Carlo runs here for different "true" specification — most especially, when we specify the model with different ratios of choice specific and individual specific coefficients. Recall that one of the findings in Keane's (1992) work was that the more restrictions which can be made in the MNP model, the better the model was able to recover the "true" parameter values.

[7]Each set of Monte Carlo trials was begun with the same seed to the random number generator. For all of these Monte Carlo simulations, we used Gauss 3.1.4 and Maxlik 3.1.3 on an IBM RS/6000.

[8]Recall that since we are using the F error covariance matrix, the true values of the error covariances are $\sigma_{12} = -.2$, $\sigma_{13} = -.5$ and $\sigma_{23} = .8$. When we compute the probabilities for the IP model, we assume that each of these error covariances are equal to 0.

[9]We justify the use of multivariate normal errors in the GEV Monte Carlo tests presented here by noting that we desired to test all of these random utility models on similar data sets. Were we to use a GEV distribution for these tests, and a multivariate normal distribution for the probit models, we would be introducing variation we could not control for in these experiments.

[10] And this expectation is confirmed by the result of the first column of Table 3.

7 References

Aldrich, J. H. and R. M. Alvarez. 1994. "Issues and the Presidential Primary Voter." *Political Behavior* 16: 289-317.

Alvarez, R. M. and J. Nagler. 1994. "A Comparison of Multinomial Probit and GEV for Estimating Models of Elections with More than Two Candidates." Paper presented at the Midwest Political Science Association Annual Meetings, Chicago, IL.

Alvarez, R. M. and J. Nagler. 1995. "Economics, Issues and the Perot Candidacy: Voter Choice in the 1992 Presidential Election." *American Journal of Political Science* forthcoming.

Amemiya, T. "The Estimation of a Simultaneous Equation Generalized Probit Model." *Econometrica* 46: 1193-1205.

Bolduc, D. "Generalized Autoregressive Errors in the Multinomial Probit Model." *Transportation Research B* 26: 155-170.

Born, R. 1990. "Surge and Decline, Negative Voting, and the Midterm Loss Phenomenon: A Simultaneous Choice Analysis." *American Journal of Political Science* 34: 615-646.

Bratley, P., B. L. Fox, and L. E. Schrage. 1983. *A Guide to Simulation.* New York: Springer-Verlag.

Bunch, D. S. 1991. "Estimability in the Multinomial Probit Model." *Transportation Research B* 25: 1-12.

Bunch, D. S. and R. Kitamura. 1989. "Multinomial Probit Estimation Revisited: Testing of New Algorithms and Evaluation of Alternative Model Specifications for Trinomial Models of Car Ownership." Transportation Research Group Research Report UCD-TRG-RR-44, University of California, Davis, CA.

Daganzo, C. 1979. *Multinomial Probit: The Theory and Its Application to Demand Forecasting.* NY: Academic Press.

Dansie, B. R. 1985. "Parameter Estimability in the Multinomial Probit Model." *Transportation Research B* 19: 526-528.

Hausman, J. A. and D. A. Wise. "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogenous Preferences." *Econometrica* 46: 403-426.

Keane, M. P. 1994. "A Computationally Practical Simulation Estimator for Panel Data." *Econometrica* 62: 95-116.

Keane, M. P. 1992. "A Note on Identification in the Multinomial Probit Model." *Journal of Business and Economic Statistics* 10: 193-200.

McFadden, D. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior." in P. Zarembka, ed., *Frontiers in Econometrics* Vol. 2. Amsterdam: North-Holland.

McFadden, D. 1989. "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration." *Econometrica* 57: 995-1026.

Rivers, D. 1988. "Heterogeneity in Models of Electoral Choice." *American Journal of Political Science* 32.